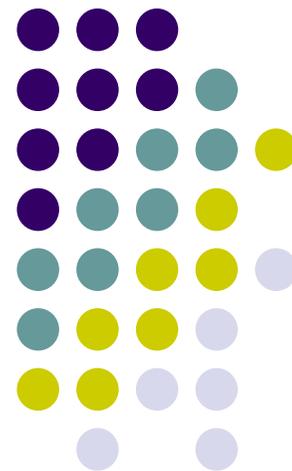


漢字構形資料庫異體字 收錄原則

——以傅斯年圖書館缺字為例

莊德明、趙苑曲
中研院資訊所文獻處理實驗室

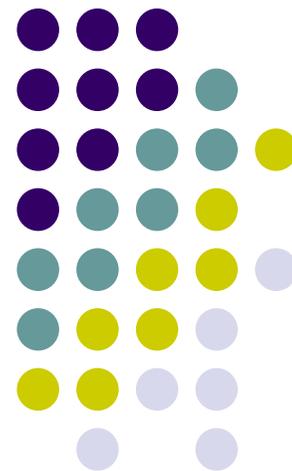




大綱

- 異體字收錄問題
- 《漢字漢語基礎》的觀點
- 〈電子古籍中的缺字問題〉的觀點
- 《漢語大字典》異寫字收錄體例
- 漢字構形資料庫異體字收錄原則
——以傅斯年圖書館缺字為例

異體字收錄問題



異體字收錄問題^{1/4}



- 由於漢字的形體不統一，圍繞著一個規範字形，會出現許多的異體字。
- 異體字的形成，一是共時的個人書寫漢字因隨意性而產生的變異；二是歷時的漢字積澱而產生的差異。
- 以異體字為部件，又會組成更多的異體字。

異體字收錄問題^{2/4}

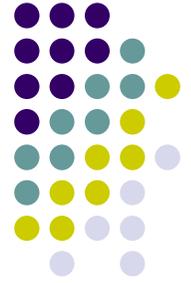
「益」的異體字



漢語大字典	益	溢	洏	洏
傳圖缺字				
異體字字典				

異體字收錄問題^{3/4}

可類推的異寫字



奇	奇 奇		
綺	綺 綺	倚	倚 倚
倚	倚 倚	寄	寄 寄

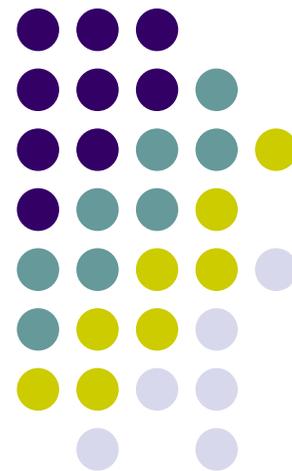
異體字收錄問題^{4/4}

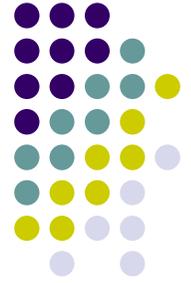


- 《廣碑別字》共收錄150個「龜」的異體字。
(詳見參考文件三)
- 漢字的異體字數量眾多，無法全數收錄，且收錄過多異體字字形，亦可能造成信息處理與使用上的不便。
- 因此針對異體字的收錄，必須訂立原則加以規範，不可照單全收。

《漢字漢語基礎》的觀點

——異寫字與異構字





異寫字與異構字

- 廣義的異體字包含了異寫字與異構字。
 - **異寫字**係指音、義相同的字，因寫法不同而造成的形體差異；這種差異並不影響字形的結構與意義。
 - **異構字**雖然在音、義上相通，但在部件選擇、部件數量、部件功能等方面，至少有一項不同，因而造成構意的差別。
- 以字形的差異是否影響字形結構與構意為區分



異寫字^{1/4}

- 異寫字是同一個字因寫法不同而造成的形體差異。如漢碑上隸書「刻」字有多種寫法：這五個「刻」字，都是左邊从「亥」，右邊从「刀」，只是「亥」和「刀」的寫法各異。

- 漢字發展初期，由於形體不固定，書寫帶有隨意性，所以異寫現象大量存在。如金文的「酉」字，可以找到數十種寫法：

- 當異寫字充當其他字的部件時，又會造成另一批異寫字。如金文「酉」作為「尊」的部件時，造成許多異寫的「尊」字：

異寫字^{2/4}



- 另外有一些字，雖然改換了部件放置的位置，但仍不影響構意，也屬於異寫字。在早期漢字裡，這種部件位置不固定的異寫字相當多。如金文的「祀」字或作「𠄎」，或作「𠄎」；金文「福」字或作「𠄎」，或作「𠄎」。
- 不同歷史時期的漢字積澱到同一共時層面上，也會因隸定方式的不同而造成異寫字。如小篆的「𠄎」字，發展到楷書出現了兩種字形：「卓」和「𠄎」，「𠄎」是將小篆字形直接描繪楷化的結果。

異寫字^{3/4}



- 由於異寫字的相互差別只是書寫方面、筆畫層次的差異，沒有構形意義上的實質差別；若加以收錄，字形雖增加，信息量卻沒有增加，反而造成漢字使用者與漢字信息處理的負擔。

異寫字^{4/4}



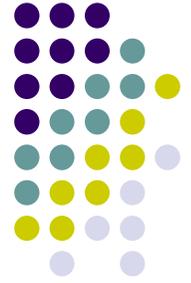
- 因此對於異寫字必須進行規範：
 1. 確定一個字形為標準字，其它異寫字則與標準字認同，不另行收錄。
 2. 對基礎部件加以規範，異寫字的形體差異，很多是在基礎部件內部產生的，所以規範基礎部件亦可減少異寫現象。如規範「酉」字的話，即可同時減少以「酉」為部件的「尊」字之異寫字。

異構字^{1/4}



- 異構字雖然音義相通，可互相置換，但在部件的選擇、部件數量、部件功能等方面，至少有一項是不同的，因此造成結構意義上的差別。
- 表音部件不同，如：「麩」，小麥屑皮也，從麥夫聲，以「夫」表音；其異構字「麩」則改以「甫」表音。
- 表義部件不同，如：「帙」，書衣也，從巾失聲，以「巾」表義；其異構字「袂」則改以「衤」（衣）表義。

異構字^{2/4}



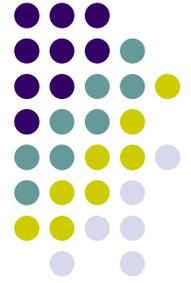
- 增加或減少部件，如「洼」和「窪」，「福」和「畐」。
- 構形方式不同，如「黥」，墨刑在面也，从黑京聲，是形聲字；其異構字「黥」改从刀，則變成从黑从刀的會意字。

異構字^{3/4}



- 由於異構字的差別，是部件及部件功能等構形屬性上的不同，因此不能將它們認同為一個字，只能說它們是記詞功能相同，但構形不同的數個字。
- 因為異構字在構形屬性上的差別，使其構意也不盡相同。在利用字形探求字的本義時，異構字可以幫助我們從更多角度去認識，因此有收錄的價值存在。

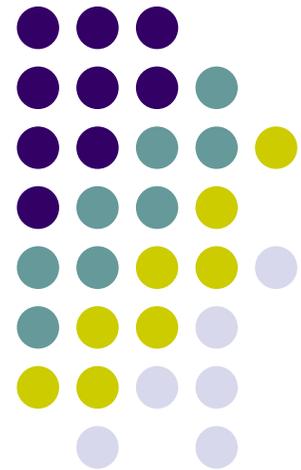
異構字^{4/4}



- 異構字雖然在社會通用的層面上來看，和異寫字一樣會造成使用者的負擔，但因其具有記錄不同構意的功能，故不能貿然取消。
- 規範異構字只能優選一個字形作為標準字使用，但其它異構字亦需保留、收錄，因為這些字在別的歷史層面上可能分別承擔著幾種不同的功能。

《電子古籍中的缺字問題》

——字形變化的表達與區分



字形變化的表達與區分^{1/5}

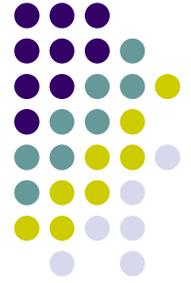


- 「字」是表達一種或一群概念的名相，是抽象的，以語意區別。如：滿、漢是表達兩種不同概念的兩個字。
- 「字形」則係以組成結構，亦即構字，作為區別。如：漢、汊是同一個字的繁簡字形。
- 以數學關係來說，字之於字形多是一對一或一對多的關係。如：「漢」字可對應到漢、汊、滙等多個字形。

字形變化的表達與區分^{2/5}



- 謝清俊老師認為電腦中缺乏字形信息，或者可說是字與字形不分，因此無法分別與處理異體字，為了在電腦中表達字與字形間的對應關係，將字形變化的表達區分如下。
- 字形往下可拆分為部件、字根、筆劃，據此可將字形的變化歸納為A.筆劃的變化、B.字根或部件的變化、C.整個字的變化三個等級。



字形變化的表達與區分^{3/5}

A. 筆劃的變化函數

A₁. 一筆劃位置改變，筆劃數和構字字根不變。如：羊→𦍋

A₂. 一筆劃尾部加勾，筆劃數和構字字根不變。如：七→𠂇

A₃. 一筆劃被另一筆劃替代，筆劃數和字根不變。

如：刃→刃、言→言

A₄. 增多一筆，筆劃數加1。如：者→耆

A₅. 減少一筆，筆劃數減1。如：德→德

A₆. 一筆劃由另二筆劃取代，筆劃數加1。如：氏→氏

A₇. 二筆劃由另一筆劃取代，筆劃數減1。如：止→止

A₈. 一群筆劃由另一群筆劃取代。如：四→𠄎

字形變化的表達與區分^{4/5}



B. 字根或部件的變化函數

B₁. 一字根由另一字根取代，而兩字根的差異只是筆劃上的變化。如：寺→寺、吉→吉

B₂. 一字根由另一字根取代，而兩字根的差異不屬於筆劃的變化。如：耻→恥

B₃. 一部件（一群字根）由另一部件取代。如：却→卻

B₄. 增多一字根。如：果→菓

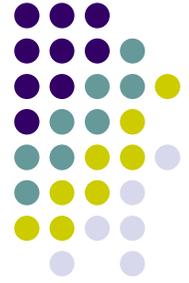
C. 整個文字構字的改變函數

C₁. 字根不變而組合改變。如：闊→潤

C₂. 因簡化而改變。如：爲→為→为

C₃. 不規則變形者。如：半→咩

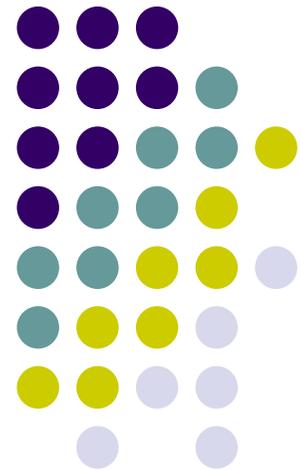
字形變化的表達與區分^{5/5}



- 上述 $A_1 \sim A_8$ 與 B_1 的變化，稱為一個字形的分毫字樣(Micro-difference)。
- $B_2 \sim B_4$ 、 $C_1 \sim C_3$ 的變化，可產生另一個字形，即異體字。

《漢語大字典》

——異寫字收錄體例





《漢語大字典》異寫字收錄體例^{1/6}

- 同時收錄以主體字為部件的字形，也收錄以異寫字為部件的字形。

朵	躲	垛	趲	蹂	搽	剝	
朶	躲	垛	垛	蹂	搽		𪗇
雋	攜	鑄	樵	僑	窩	劓	
隹	携	鑄	樵	僑	窩		𪗇

《漢語大字典》異寫字收錄體例^{2/6}



	員 鼎		
	嬋 孀 勳 勳		
	隕 隕		勳 勳 勳



《漢語大字典》異寫字收錄體例^{3/6}

𠄎	更 𠄎		
𠄎	便 𠄎	𠄎	𠄎 𠄎
𠄎	梗 𠄎	𠄎	𠄎 𠄎
𠄎	鄆 鄆	𠄎	𠄎 𠄎



《漢語大字典》異寫字收錄體例^{4/6}

- 同時收錄主體字和異寫字，但只收錄以主體字為部件組字的字形，不收錄以異寫字為部件組字的字形。

永	詠	泳	羨	咏	桼	脉	咏
𠂇						𠂇	
魚	漁	鮮	鯉	鯽	鰓	鰭	鱗
𩺰							



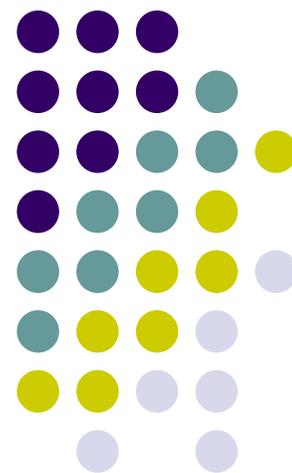
《漢語大字典》異寫字收錄體例^{5/6}

- 雖收錄因部件異寫而產生的異寫字，但並未同時收錄對應的主體字組字字形。

斤	折	析	沂	哲	誓		
						塹	螻

漢字構形資料庫異體字 收錄原則

——以傅斯年圖書館缺字為例





漢字構形資料庫異體字收錄原則^{1/6}

- 異體字的著錄原則可參考《漢字漢語基礎》：
保留異構字，規範異寫字。
- 異體字的收錄原則可參考《漢語大字典》：
異構字全收，異寫字斟酌收錄。
- 若因字源不明，難以判斷異構字，可利用〈電子古籍中的缺字問題〉的B₂-B₄、C₁-C₃的部件變化函數判斷。Ex. 「卯」與「」，部件已從「卩」變成「卩」；「導」與「」，部件從「寸」變成「小」，部件已有明顯改變，故需加以收錄。
(B2函數)



漢字構形資料庫異體字收錄原則^{2/6}

- 異寫字的收錄原則，可參考《漢語大字典》收錄體例，斟酌收錄：
 1. 因部件異寫而產生的異寫字，若《漢語大字典》常兩兩收錄，原則上可收錄。Ex.

番	幡	幡	潘	審
番	幡	幡		

漢字構形資料庫異體字收錄原則^{4/6}



- 3. 《漢語大字典》沒有出現類似的異寫字收錄體例，不收錄。
- 其他：
 1. 兩岸字形規範不同所造成的筆畫差異，不另收錄。Ex. 《漢語大字典》文本中「亮」下方部件寫作「几」，「几」和「儿」只是兩岸字形規範不同所造成的差異，因此漢字構形資料庫直接採用「亮」，不另收錄異寫字形「亮」。

漢字構形資料庫異體字收錄原則^{5/6}



- 2. 非楷體字形，如：行書、草書、圖形…等，皆不收錄。Ex.  (歌)、 (此)
- 3. 字形不清、字源不明或缺筆字，暫不收錄。Ex.  字形不清； 字源不明； 為缺筆字。



漢字構形資料庫異體字收錄原則^{6/6}

- 字形產生器：部分較單純的異寫字形，可用漢字構形資料庫部件集內的部件，加上單純橫連、直連或包含的構字符號組成構字式，則雖字形不收錄於漢字構形資料庫中，在網頁上仍能透過「字形產生器」產生字形圖片呈現。Ex.

魚	漁	鮮	鯉	鱗
𩺰			𩺰	鱗
			𩺰△里	𩺰△彡



傅斯年圖書館缺字收錄情形^{1/5}

- 缺字清單共列有429字，其中有20個重複的字形，有6字為漢字構形資料庫已收錄字形；因此實際的缺字字數為403字。

漢字構形資料庫已收錄字形

缺字編號	楷書編號	缺字編號	楷書編號	缺字編號	楷書編號
178	20560	225	71151	358	21539
伎		達		灭	
缺字編號	楷書編號	缺字編號	楷書編號	缺字編號	楷書編號
390	50802	107	17842	116	57817
遊		異		黄	



傅斯年圖書館缺字收錄情形^{2/5}

- 傅圖403個缺字中，共收錄89字；其中22字為參考〈電子古籍中的缺字問題〉部件變化函數而收錄的異構字，67個為參考《漢語大字典》收錄體例而收錄的異寫字。
- 餘314個異寫字，因《漢語大字典》無類似的收錄體例，故不收錄。



傅斯年圖書館缺字收錄情形^{3/5}

卯	寇	蕤	導	器	厚	築	惠
含	譖	竊	馱	記	覺	奮	奮
麗	降	靈	指	詣	憂		

可收錄的傅圖異構字



傅斯年圖書館缺字收錄情形^{4/5}

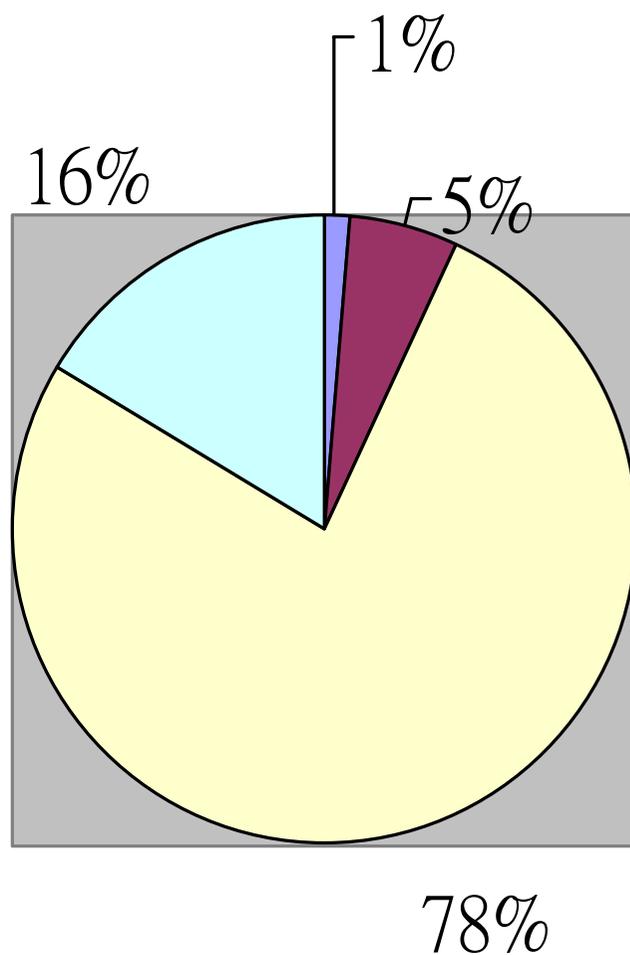
漢語大字典	斷	斷	芟	芟	紱	紱	私	私
傅圖缺字	繼		拔		袂		畝	
漢語大字典	輻	輻	懔	懔	處	處	據	據
傅圖缺字	審		潘		虎		遽	

傅圖缺字中可類推的異寫字字例

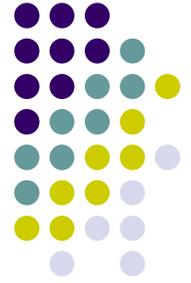


傅斯年圖書館缺字收錄情形^{5/5}

傅斯年圖書館缺字收錄情形



- 已收錄字形
- 可收錄異構字
- 不收錄異寫字
- 可收錄異寫字



參考文件

- 《漢字漢語基礎》，王寧主編，1999年，科學出版社。
- 《電子古籍中的缺字問題》，謝清俊，1996年，第一屆中國文字學會學術研討會。
- 《遠東·漢語大字典》，徐中舒主編，1991年，遠東圖書公司。